

Erfahrungen beim Aufbau eines Auswahlrahmens für Telefonstichproben in Deutschland

Gabler, Siegfried; Häder, Sabine

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Gabler, S., & Häder, S. (1999). Erfahrungen beim Aufbau eines Auswahlrahmens für Telefonstichproben in Deutschland. *ZUMA Nachrichten*, 23(44), 45-61. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-208171>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ERFAHRUNGEN BEIM AUFBAU EINES AUSWAHLRAHMENS FÜR TELEFONSTICHPROBEN IN DEUTSCHLAND

SIEGFRIED GABLER UND SABINE HÄDER

Im folgenden Bericht werden unsere Erfahrungen beim Aufbau eines Auswahlrahmens für Telefonstichproben nach dem neuentwickelten Gabler-Häder-Design vorgestellt. Weiterhin wird am Beispiel konkreter Auswahlen gezeigt, wie aus diesem Rahmen Telefonstichproben für bundesweite Befragungen konstruiert werden können.

The report presents our research findings on a frame for telephone samples using the „Gabler-Haeder“ design. Concrete samples are provided of how to draw telephone samples for nationwide surveys from this frame.

1. Ortsnetzbereiche als Systematik für den Aufbau einer Datenbank

Voraussetzung für die Ziehung einer bundesweiten Telefonstichprobe nach dem Gabler-Häder-Design (vgl. Gabler/Häder 1997; Häder/Gabler 1998; Gabler/Häder 1999) ist die Konstruktion eines adäquaten Auswahlrahmens. Für unser Design bedeutet dies, daß eine Datenbank aller mit Rufnummern besetzten Blocks in Deutschland erstellt werden muß. Diese läßt sich am effizientesten organisieren, wenn die Ortsnetzbereiche als Gliederungseinheiten verwendet werden. „Zur Adressierung der Anschlüsse von Nutzern in öffentlichen Telefonnetzen ist Deutschland in Ortsnetzbereiche eingeteilt. Jedem Ortsnetz ist eine Kennzahl (Ortsnetzkenzahl) zugeordnet. Sie beginnen mit einer Ziffer zwischen 2 und 9, der das Prefix (0) vorangestellt ist. Die Ortsnetzkenzahlen erlauben einen Rückschluß auf die geographische Lage des Anschlusses“ (Regulierungsbehörde für Telekommunikation und Post).

Ortsnetzbereiche und Gemeinden sind in der Regel nicht deckungsgleich. Dies bedeutet, daß eine Gemeinde von mehreren Ortsnetzen geschnitten werden kann wie auch, daß mehrere Gemeinden zu einem Ortsnetz gehören können. Eine Übersicht über die Überdeckung der Gemeinden mit Ortsnetzen liefert das „Ortsnetzverzeichnis Gemeinden (ONVG)“, das die

Regulierungsbehörde für Telekommunikation und Post (<http://www.regtp.de>) bereitstellt. Aus diesem kann entnommen werden, daß z.B. in Aachen Rufnummern aus den Bereichen 02404 Alsdorf Rheinland, 02407 Herzogenrath-Kohlscheid, 02408 Aachen-Kornelimünster und 0241 Aachen auftreten können. Für diese müßten dann jeweils alle besetzten Blocks extrahiert und separat gespeichert werden. Würden also die 14.275 Gemeinden (Gebietsstand 6/98)¹ als Organisationseinheiten definiert, müßten für diese Gemeinden jeweils alle besetzten Blocks aus allen zulässigen Ortsnetzbereichen berücksichtigt werden. Diese Form der Verwaltung wäre auch deshalb erheblich aufwendiger als die Organisation des Rufnummern- bzw. Blockbestandes über die Ortsnetzbereiche, da die gleichen Rufnummernblocks innerhalb eines Ortsnetzbereiches in mehreren Gemeinden auftreten können.

Folgende Quellen für Übersichten der Ortsnetzbereiche haben wir verwendet:

- Die Regulierungsbehörde stellt ein Ortsnetzbereichs-Verzeichnis (ONB-Verzeichnis) strukturiert nach dem Kriterium „Vorwahl“ (ONVV) im Internet als PDF-Datei zur Verfügung. Dieses wird regelmäßig aktualisiert. Für den Aufbau unserer Datenbank haben wir das „Verzeichnis der Ortsnetze nach Vorwahlen und Ortsnetzen sortiert“ mit Stand vom 21. August 1998 genutzt.
- Eine weitere Möglichkeit bietet die Übersicht der Kennzahlen im Telefonnetz/ISDN, FTZ 14 AB 11, Ausgabe 1997, herausgegeben und bearbeitet vom Forschungs- und Technologiezentrum der Telekom.
- Die im Internet unter der Adresse <http://set.gmd.de/~halfmann/daten/avon.txt> zur freien Verfügung bereitgestellten Vorwahldaten sind dagegen nicht vollständig, z.B. fehlen die Ortsnetzbereiche 09070, 09078, 09080, 09084, 09089, 09090, 09097 und 09099, andere Bereiche wie z.B. 033052 sind mittlerweile aufgehoben. Der für diese Datei angegebene Stand vom 1.1.1996 ist damit nicht ausreichend aktuell.

Aus dem Abgleich der genannten Quellen sowie Überprüfungen in einer aktuellen Telefonbuch CD-ROM der DeTeMedien haben wir schließlich eine Datei mit den Vorwahlen der 5201 Ortsnetzbereiche zusammengestellt².

2. Auswahl der CD-ROM

Die Registrierung der in einem Ortsnetz vergebenen Rufnummernblocks setzt eine (maschinenlesbare) Sammlung aller in diesem Ortsnetz vergebenen Rufnummern voraus. Eine

¹ Wir danken Frau Heckel (BIK ■ Aschpurwis + Behrens GmbH) für diese Angabe.

² Einer dieser Ortsnetzbereiche (08329 Riezern) überdeckt das Kleinwalsertal, dessen Bevölkerung zu Österreich gehört. Wenn die Grundgesamtheit nur die in Deutschland lebende Bevölkerung umfassen soll, ist dieses Ortsnetz auszuschließen.

derartige Sammlung wenigstens der eingetragenen Rufnummern stellen die Telefonbuch CD-ROMs dar.

Die für unsere Zwecke geeignetste und aktuellste Telefonbuch CD-ROM war die DT Info & Route '99, da sich von dieser die Ortsnetzbereiche relativ problemlos exportieren ließen, indem die Bereiche aufgerufen, markiert und die Suchbereiche exportiert wurden. Nach Angaben in der Hilfe-Datei dieser CD-ROM kann man angeblich nur bis zu 5.000 Datensätze auf einmal exportieren. Dies stimmt aber so nicht: Selbst bei Filtervorgaben lassen sich bis zu 10.000 Datensätze herunterladen. In den herausgeschriebenen Textdateien haben wir allerdings eine Vielzahl von Fehlern gefunden. Wir werden diese im folgenden relativ ausführlich darstellen, da wir damit vor dem unbedarften Gebrauch von Telefonbuch CD-ROMs zur Stichprobenziehung warnen wollen.

- Mehr als 70 Ortsnetze sind auf der CD-ROM nicht bzw. nicht vollständig enthalten. Es handelt sich insbesondere um solche, deren Ortsnetzbezeichnung mit „Sch...“ oder „Sa...“ beginnen, z.B. 03663 Schleiz, 036841 Schleusingen, 03683 Schmalkalden, 05582 Sankt Andreasberg. Dadurch fehlen mehrere Gemeinden auf der CD-ROM völlig, wie z.B. – neben den oben genannten - Saßnitz (038392) oder Schifferstadt (06235) bzw. sind nur unvollständig aufgenommen (z.B. 06129 Schlangenbad – nur ein Eintrag oder 034491 Schmöln – nur Mobilfunkteilnehmer). In diesen Fällen haben wir die Vorwahlbereiche von der CD-ROM DT-Info Gold Version '98 exportiert, da deren Datensätze identisch aufgebaut sind.
- Die Ortsnetzbereiche 069 (Frankfurt am Main) und 03984 (Prenzlau) konnten nicht geschlossen exportiert werden, obwohl sie auf der CD-ROM vorhanden sind. Offenbar liegt hier ein physischer Fehler auf dem Datenträger vor. So mußten durch Filtersetzung für die einzelnen Buchstaben die Teilnehmer in Unterdateien exportiert und diese anschließend zusammengesetzt werden. Hierbei war jedoch zu beachten, daß bei der Setzung einer Wildcard nach dem Anfangsbuchstaben des Teilnehmernamens alle Namen mit Umlauten an der zweiten Stelle des Namens nicht exportiert wurden. Dies ist der Fall, obwohl die Anwendung des Fragezeichens als Wildcard angeboten wird. Damit wären bei diesem Vorgehen nur ca. 307.000 der insgesamt 320.000 Rufnummern im Ortsnetzbereich 069 exportiert worden. Da ausländische Namen häufiger an zweiter Stelle einen Umlaut aufweisen als deutsche, wären schließlich Ausländerhaushalte in der Auswahlgesamtheit und damit auch der Stichprobe unterrepräsentiert gewesen.
- Der Ortsname Dresden taucht auf der CD-ROM zu oft auf, nämlich in 174 Ortsnetzen, wie z.B. in 02104 Mettmann (23mal), in 04821 Itzehoe (23 mal) und in 05551 Northeim (19 mal). Es gibt aber in Deutschland lediglich eine Gemeinde dieses Namens (Ortsnetz 0351). In den anderen Fällen handelt es sich um Bewohner von Gemeinden in

den angegebenen Ortsnetzen, denen ein falscher Ortsname zugewiesen wurde. Dies ergab ein Rückcheck in einer DeTeMedien CD-ROM. So sind z.B. Teilnehmer in der DT-Info '99 im Vorwahlbereich 02104 mit der Postleitzahl-Angabe 40822 Dresden versehen. Dies ist aber keine für Dresden gültige Postleitzahl, sondern ein Postleitbezirk in Mettmann. Tatsächlich finden sich auch die erwähnten Teilnehmer auf der CD-ROM der DeTeMedien mit gleichen Rufnummern, gleichen Adressen und lediglich anderen Ortsnamen – nämlich Mettmann.

- Beim Export der Vorwahlbereiche wird der letzte Datensatz in der Regel verdoppelt.
- Weiterhin haben wir festgestellt, daß Mobilfunkteilnehmer, die ihre Adresse nicht in das Telefonbuch eintragen ließen, mitunter in mehreren, darunter auch falschen Ortsnetzen eingetragen sind, da sie verschiedenen Gemeinden zugeordnet werden, die mit den gleichen Ziffern der Postleitzahl beginnen. So gibt es z.B. Mobilfunkteilnehmer, die sowohl im Ortsnetzbereich 02151 Krefeld als auch 0203 Duisburg auftauchen, da die Postleitzahlen beider Orte mit 47... beginnen, tatsächlich aber in Krefeld wohnen. Bei einer Ziehung von Gemeinden - z.B. Duisburg - aus der CD-ROM könnten damit Elemente ausgewählt werden, die nicht in die Stichprobe gehören.

Es gilt nochmals zu betonen, daß diese Fehler auf derjenigen CD-ROM gefunden wurden, deren Erscheinungsdatum am kürzesten zurücklag. Gelegentlich werden aber auch verbilligte Restbestände wesentlich älterer CD-ROMs im Handel angeboten. Der Statistiker muß sich darüber im klaren sein, daß er bei der Verwendung solcher veralteten Datenbestände zusätzlich zu den auf allen CD-ROMs mehr oder weniger verbreiteten „Flüchtigkeitsfehlern“ Fehler durch die mangelnde Aktualität der Daten macht.

3. Bereinigung, Extraktion der Rufnummern und Feststellung der Blocks

Nach Korrektur bzw. Ersetzung fehlerhafter Exportdateien haben wir in einem Bereinigungsprozeß in den einzelnen Ortsnetzen jeweils folgende Rufnummern gestrichen (vgl. auch den Beitrag „Generierung von Telefonstichproben mit TelSuSa“ in diesem Heft):

- Datensätze, bei denen als Name ausschließlich der Eintrag „Fax“ bzw. „Telefax“ steht
- Datensätze mit Vorwahlnummern, die nicht der des jeweiligen Ortsnetzes entsprechen (d.h. insbesondere Fremdeinträge und Mobilfunknummern)
- Datensätze, in deren Rufnummer ein Bindestrich steht (Geschäftsanschlüsse, die zumeist ganze Blocks besetzen)
- Datensätze, die doppelt auftreten
- Datensätze mit den Rufnummern 110 und 112

- Datensätze unzulässiger Länge (Summe der Ziffern aus Ortsnetzkennzahl und Rufnummer ist zu gering oder zu hoch³).

Mit dieser Reduktion der Rufnummern wird – neben der Eliminierung offensichtlicher Fremdeinträge – eine Einschränkung auf solche Blocks angezielt, die die Einträge der Grundgesamtheit der privaten Telefonhaushalte repräsentieren. Dies bedeutet natürlich nicht, daß mit diesem Vorgehen sämtliche Geschäftsnummern eliminiert werden, da diese durchaus in Blocks liegen können, in denen auch Privatan Anschlüsse vorkommen. Letztlich muß im Rahmen der Befragung sichergestellt werden, daß es sich bei der angewählten Rufnummer tatsächlich um die eines Privathaushaltes handelt.

In den 5.201 Ortsnetzen haben wir auf diese Weise insgesamt ca. 31.031.000 Rufnummern in 572.000 Blocks extrahiert. Das entspricht bundesweit einer mittleren Eintragsdichte von 54,2 Rufnummern pro Block. Geht man davon aus, daß ca. 20% der Rufnummern nicht in die Telefonbücher eingetragen sind, ergibt sich damit bei unserem Design im Mittel eine „Nichttrefferquote“ (Komplement der Hitrate) von ca. einem Drittel der generierten Ziffernfolgen. Der Stichprobenplan kann damit im Vergleich zu anderen Generierungsverfahren als effizient bezeichnet werden. Bei dieser Einschätzung wird jedoch als Gütekriterium nicht nur die Hitrate zugrundegelegt, sondern auch die wesentliche Eigenschaft unseres Designs, gleiche Inklusionswahrscheinlichkeiten für eingetragene und nichteingetragene Rufnummern zu liefern, das letztlich die statistische Qualität sichert.

Für die einzelnen Vorwahlbereiche (darunter verstehen wir die Menge der Ortsnetze, die – nach dem Prefix 0 – mit der gleichen Ziffer $i = 2, \dots, 9$ beginnen) haben wir die folgenden Eintragsdichten ermittelt (vgl. Tabelle 1). Die Eintragsdichte liegt lediglich im Vorwahlbereich 3 deutlich unter 50 %. Bei diesem Vorwahlbereich handelt es sich um das weitgehend neu ausgebaute Telefonnetz in Ostdeutschland. Hier ist anzunehmen, daß die Blocks nicht tatsächlich weniger dicht besetzt sind als in Westdeutschland, sondern daß bei Neuanschlüssen häufig die Option des Nichteintrags gewählt wurde.

³ Der Rückcheck einer Stichprobe von Teilnehmereinträgen aus der DT-Info'99 in der CD-ROM der DeTeMedien hat z.B. ergeben, daß in Ortsnetzen mit vierstelliger Rufnummer alle dreistelligen Einträge fehlerhaft sind. Deshalb lassen wir in vierstelligen Ortsnetzen generell dreistellige Nummern nicht zu. Gleiches gilt natürlich für die vier Ortsnetze mit dreistelliger Vorwahl. Da diese fehlerhaften dreistelligen Teilnehmernummern häufig in Blocks liegen, die mit nur einer eingetragenen Rufnummer belegt sind, führt die Eliminierung derartiger Nummern zu einer Erhöhung der Eintragsdichte.

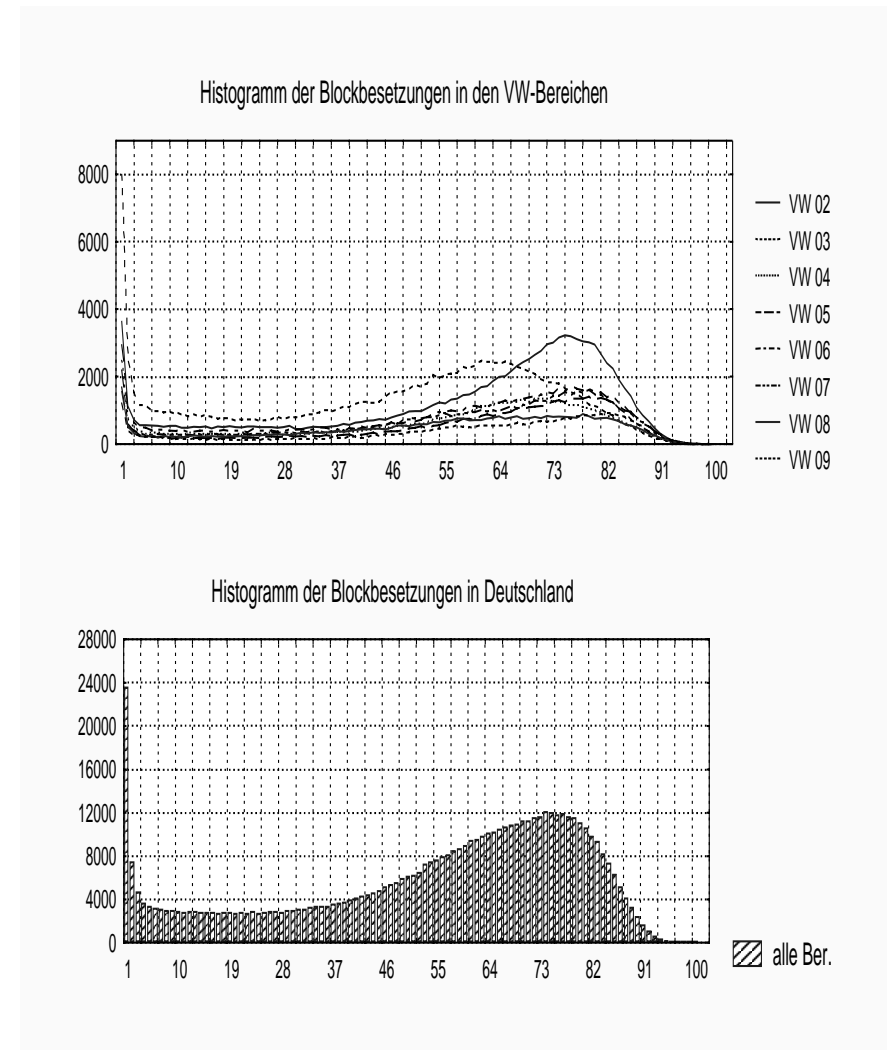
Tabelle 1: Eingetragene Rufnummern, Blocks und Eintragsdichten in den verschiedenen Vorwahlbereichen

Vorwahlbereich	Zahl der eingetragenen Rufnummern (nach Bereinigung)	Zahl der mit eingetragenen Rufnummern besetzten Blocks	Eintragsdichte
2	6.579.000	113.000	58,2
3	5.787.000	125.000	46,3
4	3.195.000	59.000	54,1
5	3.173.000	53.000	59,9
6	4.092.000	75.000	54,6
7	3.591.000	64.000	56,1
8	2.483.000	46.000	53,9
9	2.131.000	37.000	57,6

Das geringfügig abweichende Muster Ostdeutschlands in bezug auf die Blockbesetzungen ist auch aus dem Histogramm der Blockbesetzungen in den VW-Bereichen ersichtlich (vgl. Abbildung 1): Nach einem Extremum an der Stelle der mit nur einer Nummer besetzten Blocks fällt die Kurve des Vorwahlbereiches 03 wie die der anderen Bereiche auch. Es ist dann ebenfalls wieder ein Maximum bei den höher besetzten Blocks erkennbar, allerdings ist dieses auf der x-Achse weiter nach links verschoben (etwa bei 60-63 eingetragenen Nummern pro Block) als bei den anderen Bereichen (etwa bei 75-80 eingetragenen Nummern pro Block). Auffällig ist weiterhin die vergleichsweise höchste Zahl von mit nur einer Nummer besetzten Blocks, es sind hier insgesamt 6,4% aller Blocks in diesem Vorwahlbereich. In den anderen Bereichen liegt dieser Wert bei durchschnittlich 3,5%.

Interessant ist, daß das Muster der Blockbesetzungen, wie es für Deutschland insgesamt in der Abbildung 1 dargestellt ist, mit nur geringen Abweichungen in nahezu jedem Ortsnetzbereich anzutreffen ist. Deshalb sind letztlich keine wesentlichen Veränderungen in bezug auf die Hitraten zu erwarten, wenn nicht die Anteile nichteingetragener Teilnehmer von Ortsnetz zu Ortsnetz stark schwanken. Auffällig ist insbesondere, daß jeweils die Zahl der mit nur einer Nummer besetzten Blocks relativ hoch ist. Da es dafür keinen technischen Grund geben dürfte, sondern die Regulierungsbehörde für Telekommunikation und Post im Gegenteil hohe Blockbesetzungen durchsetzt, vermuten wir, daß es sich in diesen Fällen häufig um nicht korrekte Nummern handelt.

Abbildung 1: Blockbesetzungen in den Vorwahlbereichen und in Deutschland insgesamt



So haben wir Beispiele gefunden, bei denen längst ungültige, nicht mehr geschaltete Nummern nicht aus dem aktuellen Datenbestand entfernt worden sind und dadurch inzwischen eigene - leere - Blocks definieren. Weiterhin sind Sammelnummern, die auf eine Null enden, nicht immer mit einem Bindestrich versehen. Dadurch werden sie bei unserer Bereinigung auch nicht eliminiert. Sie bilden Blocks, die nicht zu unserer Gesamtheit gehören. Außerdem ist es wahrscheinlich, daß bei der Übertragung der Telefonbücher auf die CD-ROMs Fehler auftreten. Diese und sicher noch andere Gründe führen zu dem unlogischen Gipfel an der Stelle der mit nur einer eingetragenen Rufnummer besetzten Blocks.

In einem Experiment⁴ haben wir geprüft, welche Verteilung der Kontaktstati Ziffernfolgen haben, die aus „1er Blocks“ generiert worden sind. Dazu konstruierten wir eine Stichprobe vom Umfang $n=1.500$ für vier deutsche Gemeinden. Von diesen 1.500 Ziffernfolgen lagen 132 in Blocks, in denen nur jeweils eine eingetragene Nummer vorkam. In 91 Fällen, d.h. 68,9 % stellte sich heraus, daß es sich eindeutig um nicht existente Rufnummern handelte (Ansage „Kein Anschluß unter dieser Nummer“ sowie „Rufnummer geändert“), bei sechs Ziffernfolgen ertönte bei allen zehn Kontaktversuchen ein Freizeichen, so daß nicht entschieden werden konnte, ob es sich um eine geschaltete Nummer handelte. Nur in 35 Fällen waren die generierten Ziffernfolgen zweifelsfrei Telefonnummern. Die folgende Abbildung 2 veranschaulicht die Beziehung von Blockbesetzung und Ziffernfolgenstatus.

Die Punkte für die Ziffernfolgen in den relativ niedrig besetzten Blocks, insbesondere der für die 1er Blocks, liegen in der korrespondenzanalytischen Grafik auf der linken Seite der ersten Dimension (x-Achse), die bereits 94,5 Prozent der Variation in den Daten erklärt. Dort befindet sich auch der Punkt für den Ziffernfolgenstatus „Keine Rufnummer“. Blieben nun die mit nur einer eingetragenen Nummer besetzten Blocks von der Auswahl ausgeschlossen, würde sich die Hitrate erhöhen. Dabei nähme man allerdings in Kauf, daß auch tatsächlich in Blocks mit nur einer eingetragenen Nummer liegende nichteingetragene Rufnummern keine Auswahlchance hätten.

Die mit 31 bis 87 eingetragenen Rufnummern besetzten Blocks liegen dagegen in dem Quadranten, in dem sich auch der Punkt für den Status „Rufnummer“ befindet.

Ein Zusammenhang des Ziffernfolgenstatus „Freizeichen“ und der Blockbesetzung etwa derart, daß dieser Status vorrangig in niedriger besetzten Blocks auftritt und damit vermutet werden kann, daß es sich bei den betreffenden Ziffernfolgen ebenfalls um nicht-existente Nummern handelt, ist nicht erkennbar. Es ist daher notwendig, eine hohe Zahl

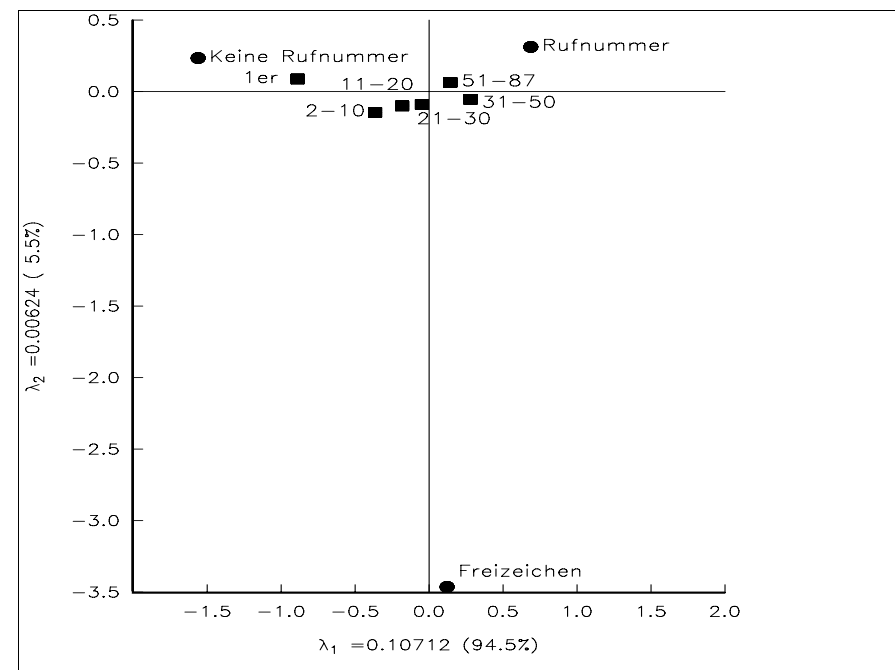
⁴ Das Experiment ist ausführlich beschrieben bei Häder/Gabler (1998) sowie bei Schneid/Stiegler (1998).

von Kontaktversuchen vorzusehen, damit nichtgeschaltete Nummern von solchen unterschieden werden können, deren Anschlußinhaber zur Zeit nicht erreichbar sind.

4. Kontrolle der Blockstabilität

Ein wesentliches Merkmal unseres Designs ist, daß es nicht direkt auf die – bei den exportfreundlichen CD-ROMs häufig nicht mehr aktuellen – Teilnehmereinträge zurückgreift, sondern diese lediglich zur Konstruktion der Blocks nutzt, aus denen schließlich Ziffernfolgen generiert werden. Hinsichtlich der Zahl der mit Rufnummern besetzten Blocks gehen wir aber von einer gewissen Stabilität über die Zeit aus: Aufgrund der mittlerweile in allen Teilen Deutschlands erreichten hohen Telefondichte sind größere Verwerfungen in der Struktur der Rufnummern in den einzelnen Ortsnetzen demnächst nicht zu erwarten.

Abbildung 2: Korrespondenz zwischen Blockbesetzung und den Stati der Ziffernfolgen



Insofern halten wir eine Aktualisierung der Blocks im Abstand von höchstens zwölf Monaten für notwendig. Dies soll an einem Beispiel demonstriert werden: Für Halle/Saale und Mannheim haben wir nach Export aus vier unterschiedlichen Telefon-CD-ROMs und anschließender Bereinigung jeweils die Anzahl der Rufnummern, der Blocks sowie die Eintragsdichten ermittelt. Es ergab sich folgendes:

Tabelle 2: Eingetragene Rufnummern, Blocks und Eintragsdichten in Halle/Saale bei verschiedenen CD-ROMs

Halle/Saale	Zahl der eingetragenen Rufnummern (nach Bereinigung)	Zahl der mit eingetragenen Rufnummern besetzten Blocks	Eintragsdichte
KlickTel '98 ⁵	84.984	2.195	38,7
Dt-Info & Route '99	84.153	2.145	39,2
DeTeMedien (August 98)	87.174	2.234	39,0
DeTeMedien (Januar 99)	85.013	2.281	37,3

Tabelle 3: Eingetragene Rufnummern, Blocks und Eintragsdichten in Mannheim bei verschiedenen CD-ROMs

Mannheim	Zahl der eingetragenen Rufnummern (nach Bereinigung)	Zahl der mit eingetragenen Rufnummern besetzten Blocks	Eintragsdichte
KlickTel '98	123.133	2.406	51,2
Dt-Info & Route '99	121.201	2.394	50,6
DeTeMedien (August 98)	122.167	2.612	46,8
DeTeMedien (Januar 99)	121.180	2.677	45,3

Während von der KlickTel '98 und der Dt-Info & Route '99 der Export von Ortsnetzbereichen mehr oder weniger komfortabel möglich ist, wird von DeTeMedien CD-ROMs nur das Herunterladen von jeweils 50 zusammenhängenden Datensätzen unterstützt. Da aber dieser Datenbestand als einziger auf dem direkten Zugriff zur Datenbank mit den Telefonbucheinträgen der Telekom beruht, ist er der aktuellste. Deshalb haben wir die

⁵ Die Daten der KlickTel sind mit denen der anderen CD-ROMs nur bedingt vergleichbar, da teilweise ein anderer Bereinigungsprozeß angewendet wurde.

Städte Mannheim und Halle/Saale zu jeweils 50 Datensätzen exportiert und schließlich die entsprechenden Dateien zusammengespielt. Dieses Verfahren ist äußerst zeitaufwendig und konnte deshalb nur exemplarisch ausgeführt werden. Die Aufschlüsse über die Qualität der Daten aus den verschiedenen CD-ROMs, die sich aus dem Vergleich der Blocks ergeben, machten diesen Aufwand aber lohnenswert.

Zunächst fällt auf, daß die Zahl der Blocks in beiden Städten auf den CD-ROMs der DeTeMedien am höchsten ist. Dies könnte insbesondere daran liegen, daß Untereinträge bei dieser CD-ROM mitexportiert werden, bei den anderen beiden Produkten jedoch nicht. Häufig haben Firmen mehrere Rufnummern, die durchaus in verschiedenen Blocks liegen. Wenn diese Rufnummern aus Untereinträgen in eigenen Blocks liegen, werden sie beim Export aus der KlickTel '98 bzw. der Dt-Info & Route '99 nicht erfaßt, obwohl sie auch bei diesen CD-ROMs als Untereinträge auftauchen. Ein Teilnehmer in Halle/Saale ist z.B. in der Dt-Info & Route '99 unter der Nummer (0345) 2029176 sowie im Zweiteintrag unter (0345) 5159330 erfaßt. Dies ist bei den DeTeMedien-CD-ROMs ebenso, allerdings wird hier der Zweiteintrag auch exportiert. Da im Block (0345) 5159300 keine andere Nummer eingetragen ist, entfällt dieser bei der Dt-Info & Route '99, nicht aber bei den DeTeMedien-CD-ROMs. Auf diese Weise erklärt sich der größte Teil der Differenzen zwischen den Blocks in den CD-ROMs der DeTeMedien und der Dt-Info & Route '99. Mehrfacheinträge sind aber ein Phänomen, das hauptsächlich für nicht-private Telefonteilnehmer typisch ist. Von daher ist der Fehler, der durch den Exportmangel der CDs hervorgerufen wird, zu vernachlässigen, wenn nicht auch nichteingetragene private Anschlüsse in diesen Blocks liegen. In einigen Fällen haben jedoch auch Privatpersonen Mehrfachanschlüsse, die in verschiedenen Blocks liegen und von denen auch mehrere ins Telefonbuch eingetragen sind. Dies ist insbesondere bei ISDN-Kunden der Fall. Hier führt der erwähnte Fehler in der Exportfunktion der Dt-Info & Route '99 sowie der KlickTel '98 eher dazu, daß den Teilnehmern mit Mehrfachanschlüssen dann keine höheren Auswahlwahrscheinlichkeiten eingeräumt werden, wenn nicht noch andere Rufnummern in solchen Blocks eingetragen sind.

Bei der Analyse der Unterschiede zwischen den besetzten Blocks, die aus der Dt-Info & Route '99 und den CD-ROMs der DeTeMedien extrahiert wurden, zeigt sich aber auch, daß in Halle seit dem Erscheinen der Dt-Info & Route '99 (Erscheinungsdatum November 1998, zu diesem Zeitpunkt nach unserer Schätzung im Mittel circa ein Jahr veraltet) durchaus neue Blocks eingerichtet worden sind. So existieren z.B. laut DeTeMedien CD-ROM (Stand August 98) die Blocks 2002500 (43 eingetragene Teilnehmer), 2002600 (38 eingetragene Teilnehmer) und 2002700 (20 eingetragene Teilnehmer), die auf der Dt-Info & Route '99 noch nicht vorkommen. Es ließen sich verschiedene Beispiele für weitere, offensichtlich neuinstallierte Blocks anführen. Teilnehmer in solchen Blocks haben bei

unserem Design aber keine Chance, ausgewählt zu werden. Deshalb muß einige Pflege auf die Aktualisierung der Blocks verwendet werden; spätestens nach zwölf Monaten sollte der Datenbestand mit einer neueren CD-ROM abgeglichen werden.

5. Stichprobenziehung

Mit der Existenz eines adäquaten Auswahlrahmens ist es nun möglich, Stichproben für Gesamtdeutschland zu konstruieren. Entsprechend unserem Design werden alle mit mindestens einem eingetragenen Teilnehmer besetzten Blocks mit der maximal möglichen Zahl von Ziffernfolgen (bei uns in der Regel 100) aufgefüllt. Aus der entstandenen Menge an Ziffernfolgen wird anschließend gezogen, bei hinreichend großem Stichprobenumfang z.B. uneingeschränkt zufällig. Denkbar ist aber auch eine vorherige Schichtung nach regionalen Merkmalen oder der Rufnummernlänge.

Um unser Design zu testen, haben wir sechs Stichproben durch uneingeschränkte Zufallsauswahl gezogen⁶. Für diese wurden die Verteilungen zweier Merkmale geprüft, von denen wir Kenntnisse über die Verteilung in der Gesamtheit haben. Zu untersuchen war

- inwieweit die Verteilungen der Ziffernfolgen über die acht Vorwahlbereiche in den Stichproben von der Verteilung aller möglichen Ziffernfolgen über die acht Vorwahlbereiche abweichen,
- inwieweit die Verteilungen der Ziffernfolgenlänge in den Stichproben von der Verteilung der Ziffernfolgenlänge aller möglichen Ziffernfolgen abweichen.

In allen Stichproben sind die Abweichungen der Verteilung der Ziffernfolgen auf die Vorwahlbereiche zur Verteilung aller möglichen Ziffernfolgen in der Gesamtheit gering, d.h. die Stichproben bilden die regionale Streuung der Rufnummernblocks jeweils gut ab. χ^2 -Anpassungstests ergaben, daß sich die Stichprobenverteilungen bei einer Irrtumswahrscheinlichkeit von 5% in fünf der sechs Stichproben nicht signifikant von denen der Gesamtheit unterscheiden.

Bei dem zweiten zu untersuchenden Merkmal handelt es sich um die Ziffernfolgen- und damit auch die Rufnummernlänge⁷. Es existieren in Deutschland 3- bis 8-stellige Rufnummern. 3-stellige Nummern wurden früher häufig in sehr kleinen Orten vergeben. Sie haben eine höhere Eintragsdichte als Rufnummern mit größerer Länge.

⁶ Die Stichprobengrößen $n=12058$ und $n=4690$ haben wir gewählt, um den direkten Vergleich zu einem weiter unten beschriebenen Design für Telefonstichproben zu ermöglichen.

⁷ Die eingetragenen und nichteingetragenen Rufnummern sind Teilmengen der Ziffernfolgen.

Tabelle 4: Relative Häufigkeiten (in Prozent) der Ziffernfolgen nach Vorwahlbereichen in sechs Stichproben und der Gesamtheit, Ergebnisse eines χ^2 -Anpassungstests

Vorwahlbereich	Stichprobe 1	Stichprobe 2	Stichprobe 3	Stichprobe 4	Stichprobe 5	Stichprobe 6	Gesamtheit
n=	4.690	4.690	4.690	12.058	12.058	2.058	
2	19,49	19,74	19,08	20,08	20,29	20,06	19,81
3	22,84	21,83	21,32	21,51	21,33	21,58	21,87
4	11,43	10,55	10,47	10,39	10,57	10,67	10,37
5	9,06	8,81	9,72	9,25	9,30	9,32	9,30
6	12,56	12,94	13,24	13,42	12,75	13,05	13,07
7	10,96	11,75	11,62	10,89	11,29	11,17	11,15
8	8,16	7,95	8,57	7,92	7,86	8,15	8,06
9	5,50	6,42	5,97	6,54	6,62	6,00	6,38
gesamt	100	100	100	100	100	100	100
Prüfgröße	14,44	3,04	6,63	3,84	6,30	4,68	-
p	0,04	0,88	0,47	0,80	0,51	0,70	-

In Langeneß-Hallig (Ortsnetzkennzahl 04684) und in Forbach-Hundsbach (Ortsnetzkennzahl 07220) kommen z.B. gemäß der DT-Info & Route '99 ausschließlich 3-stellige Nummern vor⁸. 8-stellige Nummern gibt es dagegen vorrangig in den Großstädten mit einer 3-stelligen Vorwahl (Berlin, Frankfurt, Hamburg, München). Neben der Beziehung zwischen Region und Rufnummernlänge läßt sich auch eine zwischen der Stellenzahl der Rufnummer und dem Alter der Anschlußinhaber vermuten. So sind ISDN-Anschlüsse in der Regel höherstellig und werden seltener von älteren Menschen beantragt, bei 3-stelligen Nummern sind dagegen kaum jüngere Anschlußinhaber zu erwarten. Würde also die Verteilung der Rufnummernlänge in der Stichprobe von der in der Gesamtheit abweichen, wäre mit einem systematischen Fehler zu rechnen. Da wir aber die Verteilung der Länge aller Rufnummern – das heißt aller geschalteten Ziffernfolgen - nicht kennen (wir haben keine Kenntnis über die Stellenzahlen der nichteingetragenen Nummern), untersuchten wir die der generierten Ziffernfolgen im Vergleich zur Verteilung der Stellenzahl aller nach unserem Design konstruierbaren Ziffernfolgen.

⁸ In Forbach-Hundsbach gibt es darüber hinaus noch zwei Rufnummern mit Bindestrich, die bei unserer Bereinigung eliminiert worden sind.

Auch hinsichtlich der Repräsentation der Ziffernfolgenlänge in den Stichproben ist unser Ergebnis sehr zufriedenstellend. In allen Stichproben kann bei einer Irrtumswahrscheinlichkeit von 5% davon ausgegangen werden, daß sich die Verteilung der Ziffernfolgenlänge in den Stichproben nicht von der in der Gesamtheit unterscheidet.

Tabelle 5: Relative Häufigkeiten (in Prozent) der Länge der Ziffernfolgen in sechs Stichproben und der Gesamtheit, Ergebnisse eines χ^2 -Anpassungstests

Ziffernfolgenlänge	Stichprobe 1	Stichprobe 2	Stichprobe 3	Stichprobe 4	Stichprobe 5	Stichprobe 6	Gesamtheit
n=	4.690	4.690	4.690	12.058	12.058	12.058	-
3 stellig	2,11	2,13	2,09	2,30	2,32	2,24	2,22
4 stellig	18,38	20,47	20,28	20,17	19,95	19,88	19,87
5 stellig	30,64	30,19	30,38	29,60	29,48	29,26	29,75
6 stellig	29,83	29,28	30,02	30,02	30,33	30,10	30,14
7 stellig	16,67	15,65	14,99	15,67	15,40	15,97	15,66
8 stellig	2,37	2,28	2,24	2,24	2,52	2,55	2,35
gesamt	100	100	100	100	100	100	100
Prüfgröße	9,98	2,58	3,00	1,65	3,05	3,90	-
p	0,08	0,77	0,70	0,90	0,69	0,56	-

Daß dies durchaus nicht bei allen Designs für Telefonstichproben so ist, wollen wir anhand des folgenden Beispiels belegen.

6. Bewertung eines alternativen Vorschlags für die Ziehung von Telefonstichproben: das Ziffernfolgen-Design

Das in den letzten Jahren aufgetretene Problem des sich ständig erhöhenden Anteils nicht in die Telefonbücher eingetragener Teilnehmer hat zu verschiedenen neuen Vorschlägen für die Telefonstichprobenziehung geführt (vgl. z.B. von der Heyde 1998, Gabler/Häder 1998, Rösch 1998). Unter anderem wird auch die folgende Idee diskutiert, die wir als Ziffernfolgen-Design bezeichnen wollen: Zunächst wird ein Auswahlrahmen mit der Menge aller eingetragenen Telefonnummern geschaffen. Aus diesem werden dann alle Rufnummern ausgewählt, in denen eine bestimmte, zufällig gebildete Folge von vier Ziffern auftritt. So ist es z.B. denkbar, daß die Ziffernfolge 1234 generiert wird. Es würden dann alle diejenigen Nummern ausgewählt, in denen diese Ziffernfolge auftritt, also

z.B. 1234777, 341234, 7812345 und so fort. Bei diesen Rufnummern würden anschließend die letzten beiden Ziffern zufällig ersetzt und die entstehenden Ziffernfolgen schließlich ausgewählt. Von diesem Vorgehen erhofft man sich eine gute Repräsentation eingetragener und nichteingetragener Telefonhaushalte.

Abgesehen davon, daß das Design kritikwürdig ist, da die Inklusionswahrscheinlichkeiten für die Rufnummern nicht gleich (für dreistellige Nummern sind sie z.B. Null) sind, haben wir uns dafür interessiert, wie die regionale Verteilung derartig konstruierter Stichproben aussieht. Dafür haben wir aus der Menge aller in Frage kommenden vierstelligen Ziffernfolgen vier ausgewählt (1234, 4321, 7878 und 9999) und für diese die Häufigkeiten des Auftretens innerhalb der Menge der eingetragenen Rufnummern in allen Ortsnetzbereichen ermittelt. Es stellte sich heraus, daß diese Ziffernfolgen durchaus nicht gleichhäufig innerhalb von Rufnummern auftreten:

Ziffernfolge	1234	4321	7878	9999
Häufigkeit	12.057	8.392	5.787	4.689

Auch für die ausgewählten Rufnummernblocks (die letzten beiden Ziffern werden ja laut Design zufällig ersetzt, was aber weder an der Zugehörigkeit zu einem Vorwahlbereich noch an der Rufnummernlänge etwas ändert) haben wir die oben beschriebenen Tests (vgl. Tabelle 4 und 5) ausgeführt.

Die Ergebnisse der χ^2 -Tests zeigen, daß in den vier Stichproben keine gute regionale Abbildung der Gesamtheit erreicht werden konnte, in jedem Fall wird die Hypothese verworfen, die Stichprobenverteilung und die Verteilung der Rufnummern auf die Vorwahlbereiche in der Gesamtheit seien gleich.

Ebenso verhält es sich mit der Verteilung der Ziffernfolgenlängen in allen Stichproben im Vergleich zur Gesamtheit. Es muß daher befürchtet werden, daß die Stichproben mindestens in regionalen Merkmalen, der Altersverteilung und den damit zusammenhängenden Merkmalen verzerrt wären. Deshalb ist das Ziffernfolgen-Design für die Anwendung in Deutschland ungeeignet.

7. Zusammenfassung

Für die Ziehung von bundesweiten Telefonstichproben ist der Aufbau eines adäquaten Auswahlrahmens von entscheidender Bedeutung. Hierzu muß sehr viel Sorgfalt aufgewendet werden, da die im Handel befindlichen exportfreundlichen Telefonbuch CD-ROMs offenbar Qualitätsmängel aufweisen.

Tabelle 6: Relative Häufigkeiten (in Prozent) der Ziffernfolgen nach Vorwahlbereichen in vier Stichproben nach dem Ziffernfolgen-Design und der Gesamtheit, Ergebnisse eines χ^2 -Anpassungstests

Vorwahl- bereiche	Stichprobe 1 „1234“	Stichprobe 2 „4321“	Stichprobe 3 „7878“	Stichprobe 4 „9999“	Gesamtheit
n=	12.057	8.392	5.787	4.689	-
2	18,2	23,9	24,9	12,5	19,81
3	24,6	28,6	13,3	9,6	21,87
4	8,9	7,5	15,4	13,5	10,37
5	9,6	9,3	7,4	17,9	9,30
6	10,4	12,0	16,0	18,2	13,07
7	8,8	8,9	12,7	6,2	11,15
8	13,1	6,1	5,8	12,9	8,06
9	6,4	3,8	4,4	9,2	6,38
gesamt	100	100	100	100	100
Prüfgröße	593,78	486,0	557,8	1247,2	-
p	0	0	0	0	-

Tabelle 7: Relative Häufigkeiten (in Prozent) der Länge der Ziffernfolgen in vier Stichproben nach dem Ziffernfolgen-Design und der Gesamtheit, Ergebnisse eines χ^2 -Anpassungstests

Ziffernfol- gen-länge	Stichprobe 1 „1234“	Stichprobe 2 „4321“	Stichprobe 3 „7878“	Stichprobe 4 „9999“	Gesamtheit
n=	12.057	8.392	5.787	4.689	-
3-stellig	0	0	0	0	2,22
4-stellig	19,35	10,95	14,67	2,79	19,87
5-stellig	27,56	32,47	24,83	16,76	29,75
6-stellig	27,88	36,64	42,32	50,05	30,14
7-stellig	24,69	19,55	14,05	14,33	15,66
8-stellig	0,51	0,38	4,13	16,06	2,35
gesamt	100	100	100	100	100
Prüfgröße	1110	881	627	5430	-
p	0	0	0	0	-

Der Datenbestand - die mit eingetragenen Nummern besetzten Blocks in den einzelnen Ortsnetzbereichen - sollte mindestens einmal jährlich aktualisiert werden. Dabei ist besondere Aufmerksamkeit auf die mit nur einer eingetragenen Nummer belegten Blocks zu legen, da es sich hierbei häufig um fehlerhafte Angaben, d.h. nicht existente Nummern handelt.

Mit uneingeschränkter Zufallsauswahl aus der Menge aller prinzipiell als Rufnummern möglichen Ziffernfolgen konnten in Beispielstichproben nach unserem Design gute regionale Streuungen und adäquate Abbildungen der Ziffernfolgenlängen erreicht werden. Dies traf auf das Ziffernfolgen-Design für die Stichprobenziehung in Deutschland nicht zu, das u.a. deshalb nicht empfohlen werden kann.

Literatur

Gabler, S./Häder, S., 1997: Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. ZUMA-Nachrichten 41: 7-18.

Gabler, S./Häder, S., 1998: Probleme bei der Anwendung von RLD-Verfahren. S. 58-68 in: S. Gabler/S. Häder/ J. Hoffmeyer-Zlotnik. (Hrsg.), Telefonstichproben in Deutschland. Opladen: Westdeutscher Verlag.

Gabler, S./Häder, S., 1999: Generierung von Telefonstichproben mit TelSuSa. ZUMA-Nachrichten 44.

Häder, S./Gabler, S., 1998: Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. S. 69-88 in: S. Gabler/S. Häder/ J. Hoffmeyer-Zlotnik. (Hrsg.), Telefonstichproben in Deutschland. Opladen: Westdeutscher Verlag.

Rösch, G., 1998: Maßnahmen zur Reduktion von Stichprobenfehlern bei telefonischen Bevölkerungsumfragen. S. 101-119 in: S. Gabler/S. Häder/ J. Hoffmeyer-Zlotnik. (Hrsg.), Telefonstichproben in Deutschland. Opladen: Westdeutscher Verlag.

Schneid, M./Stiegler, A., 1998: Feldbericht Telefonumfrage "Projekt Hitrate". ZUMA Technischer Bericht 98/03.

von der Heyde, Ch., 1998: InfraScope – Der Telefonbus von Infratest Burke. S. 133-140 in: S. Gabler/S. Häder/ J. Hoffmeyer-Zlotnik. (Hrsg.), Telefonstichproben in Deutschland. Opladen: Westdeutscher Verlag.